Solution 1: Multiclass Classification with 0-1-Loss

(a) As seen in the 0-1-Loss presentation, slide 2, the discrete classifier that minimizes the risk $h^*(\mathbf{x})$ (the Bayes optimal classifier) is:

$$h^{*}(\mathbf{x}) = \arg\max_{l \in \mathcal{Y}} \underbrace{\mathbb{P}(y = l \mid \mathbf{x} = \mathbf{x})}_{\sim \text{Unif}\{1, \dots, x\}}$$

$$= \arg\max \frac{1}{x} \cdot \mathbb{1}_{[1 \le l \le x]}$$
(1)

As the distribution of y given x is uniform , any value between 1 and x is optimal.

$$h^*(\mathbf{x}) = \{1, \dots, x\} \tag{2}$$

(b) The Bayes risk for the 0-1-loss, also known as the Bayes error rate, is defined as :

$$\mathcal{R}^* = 1 - \mathbb{E}_x \left[\max_{l \in \mathcal{Y}} \mathbb{P}(y = l \mid \mathbf{x} = \mathbf{x}) \right]$$

$$= 1 - \mathbb{E}_x \left[\frac{1}{x} \right]$$

$$p_x \sim \text{Unif}\{1, \dots, 10\}$$

$$= 1 - \sum_{x=1}^{10} \frac{1}{x} \frac{1}{10}$$

$$\lim_{t \to \infty} 1 - \frac{7381}{25200}$$
(3)

An alternative solution to the problem can be derived from the risk definition:

$$\mathcal{R}^* = \sum_{X} \sum_{Y} L(l, h^*(x)) \ \mathbb{P}(y = l \mid \mathbf{x} = \mathbf{x}) \ \mathbb{P}(x)$$

$$= \sum_{i=1}^{10} \frac{1}{10} \sum_{j=1}^{i} \mathbb{I}_{\{j \neq h^*(i)\}} \frac{1}{i}$$
(4)

As exactly only on value of $\{1, \ldots, x\}$ will be chosen for $h^*(x)$, the indicator function will be 1 for all values except for the chosen one. Therefore:

$$\mathcal{R}^* = \sum_{i=1}^{10} \frac{1}{10} \sum_{j=1}^{i} \mathbb{I}_{\{j \neq h^*(i)\}} \frac{1}{i}$$

$$= \sum_{i=1}^{10} \frac{1}{10} \left(1 - \frac{1}{i} \right)$$

$$= \sum_{i=1}^{10} \frac{1}{10} - \frac{1}{10} \sum_{i=1}^{10} \frac{1}{i}$$

$$\stackrel{hint}{=} 1 - \frac{7381}{25200}$$
(5)

(c) The point-wise optimizer for the 0-1 loss over all discrete classifiers $h^*(\mathbf{x})$ is:

$$h^*(\mathbf{x}) = \arg\max_{l \in \mathcal{V}} \ \mathbb{P}(y = l \mid \mathbf{x} = \mathbf{x})$$
 (6)

The optimal constant model can be obtained by forgetting the conditioning on x, leading to:

$$\bar{h}(x) = \arg\max_{l \in \mathcal{V}} \quad \mathbb{P}(y = l)$$
 (7)

Using the law of total probability:

$$\begin{split} \bar{h}(x) &= \arg\max_{l \in \mathcal{Y}} \sum_{x=1}^{10} \mathbb{P}(y = l \mid \mathbf{x} = \mathbf{x}) \cdot \mathbb{P}(\mathbf{x} = \mathbf{x}) \\ &= \arg\max_{l \in \mathcal{Y}} \sum_{x=1}^{10} \frac{1}{x} \cdot \mathbb{I}_{[1 \le l \le x]} \cdot \frac{1}{10} \\ &= \arg\max_{l \in \mathcal{Y}} \begin{cases} \frac{7381}{25200}, & l = 1\\ \frac{7381}{25200} - \frac{1}{10}, & l = 2\\ \frac{7381}{25200} - \frac{1}{10} - \frac{1}{20}, & l = 3\\ \vdots & \vdots\\ \frac{7381}{25200} - \sum_{z=1}^{l-1} \frac{1}{10 \cdot z}, & l = 10 \end{cases} \end{split}$$
(8)

As the probability is monotonically decreasing with l, we can conclude that the optimal constant model is:

$$\bar{h}(x) = 1 \tag{9}$$

(d) The Risk is calculated by:

$$\mathbb{R}_{L}(\bar{h}) = 1 - \max \mathbb{P}(y = l)$$

$$= 1 - \mathbb{P}(y = 1)$$

$$= 1 - \frac{7381}{25200}$$